DOCUMENT RESUME

ED 209 258 TM 810 712

AUTHOR Haenn, Joseph F.
TITLE A Practitioner's Guide to Functional Level Testing.
INSTITUTION National Testing Service, Inc., Durham, N.C.
PUB DATE Apr 81
NOTE 96p.; Paper presented at the Annual Meeting of the American Educational Research Association (65th, Los Angeles, CA, April 13-17, 1981).

EDRS PRICE MF01/PC04 Plus Postage.
DESCRIPTORS *Achievement Tests; *Difficulty Level; Scores; *Scoring; *Testing; *Testing Problems; Test Interpretation; Test Items; *Test Reliability; Test Validity
IDENTIFIERS *Out of Level Testing

ABSTRACT

Procedures for conducting functional level testing have been available for use by practitioners for some time. However, the Title I Evaluation and Reporting System (TIERS), developed in response to the educational amendments of 1974 to the Elementary and Secondary Education Act (ESEA), has provided the impetus for widespread adoption of this testing practice. The following policy strongly implies its usage: All project evaluations should use measures of educational performance appropriate to the curriculum and to the students served. Such appropriateness is defined in terms of: (1) reliability and validity (the content of the test should reflect the content of the project's instruction, and the pretest and posttest should measure the same abilities); and (2) difficulty level (the relationship between the difficulty level of the test and the functional or performance level of the students should be considered). There are four major objectives to this document. They are to present several general concepts which relate to the practice of functional level testing, discuss conditions under which functional level testing is appropriate, present the basis of the controversy for and against the practice of functional level testing, and describe procedures for conducting functional level testing and converting test scores to in-level normed scores. (Author/GK)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED209258

A PRACTITIONER'S GUIDE TO FUNCTIONAL LEVEL TESTING

JOSEPH F. HAENN
NTS RESEARCH CORPORATION
2634 CHAPEL HILL BOULEVARD
DURHAM, NORTH CAROLINA 27707

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.
☒ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

J. F. Haenn

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

APRIL, 1981

TM 810 712

NTS Research Corporation
2634 Chapel Hill Blvd. Durham, North Carolina 27707

A PRACTITIONER'S GUIDE TO FUNCTIONAL LEVEL TESTING[1]

Joseph F. Haenn
NTS Research Corporation

Procedures for conducting functional level testing have been available for use by practitioner's for some time. However, the Title I Evaluation and Reporting System (TIERS), developed in response to the educational amendments of 1974 to the Elementary and Secondary Education Act (ESEA), has provided the impetus for widespread adoption of this testing practice. Chapter 10 of the Policy Manual, which provides the guidelines for evaluation using TIERS, does not specifically mention functional level testing but strongly implies its usage:

> All project evaluations should use measures of educational performance appropriate to the curriculum and to the students served. Appropriateness is defined in terms of:
>
> - Reliability and validity - The test should meet the technical standards of reliability and validity described earlier.... Particularly, the content of the test should reflect the content of the project's instruction, and the pretest and posttest should measure the same abilities.
>
> - Difficulty level - The relationship between the difficulty level of the test and the functional or performance level of the students should be considered. Students should not generally be administered tests that are far too easy or far too difficult; rather, a test of reasonable difficulty should be administered.
>
> - .... (USED, 1980, p. 10).

There are four major objectives to this document. They are to:

1. Present several general concepts which relate to the practice of Functional Level Testing.
2. Discuss conditions under which Functional Level Testing is appropriate.
3. Present the basis of the controversy for and against the practice of Functional Level Testing.

---

[1] Paper presented at AERA Annual Meeting, Los Angeles, CA. April, 1981.

4. Describe the procedures for conducting Functional Level Testing and converting test scores to in-level normed scores.

The remainder of the document is devoted to sections which address each of these objectives.

General Concepts

There are several general concepts which relate to the practice of Functional Level Testing. Before presenting these concepts, it is important to present some terminology which will provide a basis for these discussions. Figure 1 presents those terms which are most commonly used. A discussion of some of these terms is first necessary.

Test publishers generally develop several contiguous levels of tests which are nominally recommended for a particular grade (or age) or range of grades (ages). Administering the test level which is recommended by the publisher for a particular grade or age is referred to as in-level testing. Synonymous terms for in-level testing include on-level testing, at-level testing, grade-level testing, and age-level testing. Out-of-level testing refers to the practice of administering a test level at a grade or age other than that nominally recommended by the publisher. Synonymous terms for out-of-level testing include below-level testing and above-level testing. For example, a particular test level may be designed for third grade students. If that test level were to be administered to third graders, this would be called in-level testing, or one of the other synonymous terms. If the same third grade level of the test were administered to another grade level, such as the fifth grade, this practice would be referred to as out-of-level testing, or one of the other synonymous terms.

Within a given grade or age level, or even within a given classroom, one test level may not be appropriate for all of the students who are to be tested. Under these circumstances, it is advisable to test students at their functional

Ceiling Effect: An effect which can result from the administration of a test which is too easy causing extremely high scores and careless mistakes. A ceiling effect can restrict the potential for detecting actual student gains.

Chance Score: The score which can be expected to be obtained by guessing (chance) alone. A chance score is defined as the number of items divided by the number of alternatives.

Floor Effect: An effect which can result from the administration of a test which is too difficult causing low scores near the chance score level. A floor effect can restrict the potential for detecting actual student gains.

Functional Level Testing: Testing students with a test appropriate for their current level of instructional functioning rather than with a test designed for their current grade placement. Functional Level Testing can involve in-level testing for some students and out-of-level testing for other students.

Individualized Testing: A special form of Functional Level Testing in which each individual pupil is tested at his/her functional level.

In-Level Testing: Administration of the test level which is recommended for testing students of a given grade placement.

Instructional Level Testing: A synonymous term for Functional Level Testing.

Off-Level Testing: A synonymous term for Out-of-Level Testing.

On-Level Testing: A synonymous term for In-Level Testing.

Out-of-Level Testing: Administration of a test level which is not recommended by the publisher for testing students of a given grade placement. Usually an out-of-level test is chosen to be appropriate to a student's functional level.

Figure 1. Terminology

or instructional level. Functional Level Testing refers to testing a student at his or her functional level of ability. Instructional Level Testing is quite similar and refers to testing a student according to the level of instruction that that particular student is receiving in the classroom. Of course, when a group of students is being tested, each particular student does not have to be tested at the same instructional or functional level. Those students who are at age or grade level would be tested in-level. Those students above or below their grade or age level in terms of instructional level or functional ability would be tested with out-of-level tests. This practice of testing students at their functional or instructional level regardless of whether it requires in-level or out-of-level tests is called Functional Level Testing or Instructional Level Testing.

Chance Scores

On a multiple choice test or similar test on which a student can respond without knowing the correct answer, it is possible for a student to get some items correct by guessing alone. The pure chance level of a test is determined by dividing the number of items (N) by the number of alternatives per item (A) (Gulliksen, 1950; Roberts, 1978a). Mathematically, this is expressed as:

$$\text{Test Chance Score} = \frac{N}{A}$$

For example, on a sixty item four-choice, multiple choice test the chance score would be fifteen raw score points (60/4). On a thirty item true/false test the chance score would also be fifteen raw score points (30/2). For tests having items with different numbers of alternatives, the chance score should be computed separately for each type of item and then these part chance scores are added together.

The chance scores for many of the available subtests in reading and mathematics which are commonly used with children in grades 2-6 are given in

Appendix A. These tables should be examined in some detail when considering the use of a given level for a particular group of educationally disadvantaged students. For example, for those tests with Total Reading scores, the chance score at the lowest designated level for in-level administration between grades 2 and 6 ranges from the 2nd to the 39th percentile. For those tests with Total Mathematics scores, the chance score at the lowest designated level for in-level administration between grades 2 and 6 ranges from below the 1st percentile to the 39th percentile. Therefore, even though students may be functioning at the chance level they may still receive a normative score as high as the 39th percentile according to the table of norms.

Because of this phenomenon, special attention should be directed towards what can happen if test levels are changed from pretest to posttest. For example, on the reading comprehension subtest of one commonly used test, if the level which is nominally recommended for the beginning of the second grade is administered as the pretest and the next higher level, which is nominally recommended for the end of the second grade, is administered as the posttest student scores could jump from the 21st percentile to the 35th percentile even though they would be scoring at the raw score chance level on each test level. On another test, by administering the level of the Total Mathematics subtest, which is nominally recommended for the beginning of grade four, as the pretest and the next higher level, which is nominally recommended for the end of grade four, as the posttest, student scores at the raw score chance level would rise from the 1st percentile to the 7th percentile. Therefore, the raw score chance levels and their corresponding grade equivalents and percentiles should be carefully examined for the test and grade(s) of interest.

Wick and Ward (1977) examined the correlations between chance scores on adjacent levels for reading comprehension, reading vocabulary, and mathematics

subtests. They determined that the average correlation between these adjacent levels on chance scores is about 0.96. Therefore, scores at the chance level cannot be easily detected by simply observing what happens from administration of two different test levels. Two additional studies have examined the issue of chance score responding and present additional procedures for detecting when these effects occur (Burton, 1979; Stamman, n.d.).

Floor Effects

A floor effect implies that a large proportion of the students are scoring at the minimal score possible on the test. This minimal score is usually considerably greater than simply no correct items except on open-ended response items, such as completion or fill-in-the-blank items. On closed-response items the minimal score is usually indicated as a chance score.

On tests where a large number of students score at or around the chance level, very little useful information is obtained. The general rule for determining whether a floor effect is present for a group of students is obtained by dividing the number of items (N) by one less than the number of alternatives per item (A - 1) (Roberts, 1978a). Mathematically, this is expressed as:

$$\text{General Rule for determining the presence of a floor effect for a group mean} = \frac{N}{A-1}$$

For four-choice items, one-third of the maximum raw score for a group of students defines the situation where a large number of students are scoring at or below the chance score level. For five-choice items, the general rule would be one-quarter of the maximum raw score. As the number of choices per item is reduced, this general rule becomes less applicable. For instance, for three-choice items the general rule predicts that a large number of students would be scoring at or below the chance score if the mean score for the group is one-half

of the maximum raw score. This general rule cannot be applied on true/false items since the general rule predicts that chance level effects are occurring when the mean score is 100% of the items maximum raw score, or a perfect score.

An alternative rule uses the raw score mean and median. If the average raw score mean of the respondent group is at least one-third of the standard deviation higher than the median raw score, then the chances are very good that these students have "bottomed-out" on this level (Roberts, 1978a). The floor levels which exist for commonly used standardized norm-referenced achievement tests are presented by Roberts (1978b). The responses illustrated in Figure 2 demonstrate a floor effect.

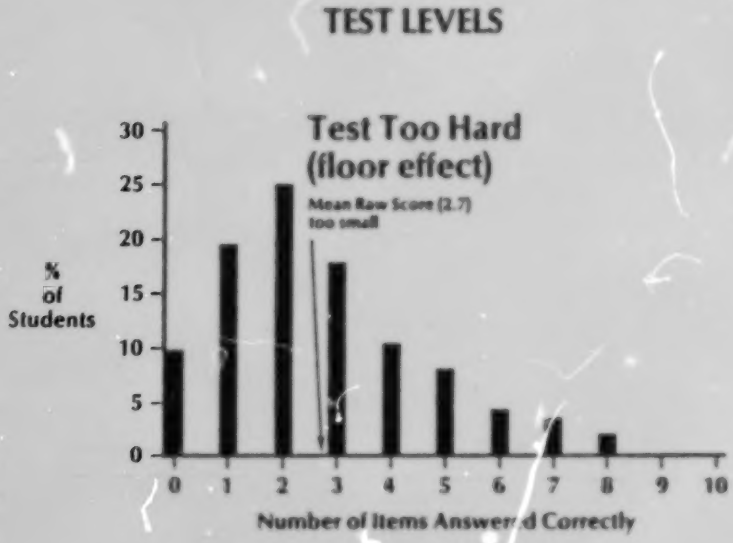


Figure 2. The floor effect which results from administering a test which is too difficult

Ceiling Effects

If students find that items are too easy they may become inattentive and make careless mistakes. A general rule of thumb to determine whether a test is too easy and a ceiling effect has been encountered for a group of students is to subtract the chance score from the number of items (Roberts, 1978a).

Mathematically, this is expressed as:

$$\text{General Rule for determining the presence of a ceiling effect for a group mean} = N - \frac{N}{A}$$

If on a test with four-choice, multiple choice items, the mean of the group exceeds a score equivalent to three-quarters or more of the raw score items correct, a "ceiling effect" may have been encountered.

A ceiling effect implies that a large proportion of the students are reaching the maximum possible score on the test. This maximum score may be less than the total number of items due to carelessness or misleading item structure. Another check to determine whether a ceiling effect has been encountered for a group of students is to compare the mean of the respondent group with their median score (Roberts, 1978a). If the mean is at least one-third of a standard deviation lower than the median, then the chances are very good that the students have "topped-out" on this test level. Roberts (1978b) reports the ceiling levels for the most commonly used tests and indicates that several of these test ceilings are quite low. The responses illustrated in Figure 3 demonstrate a ceiling effect.

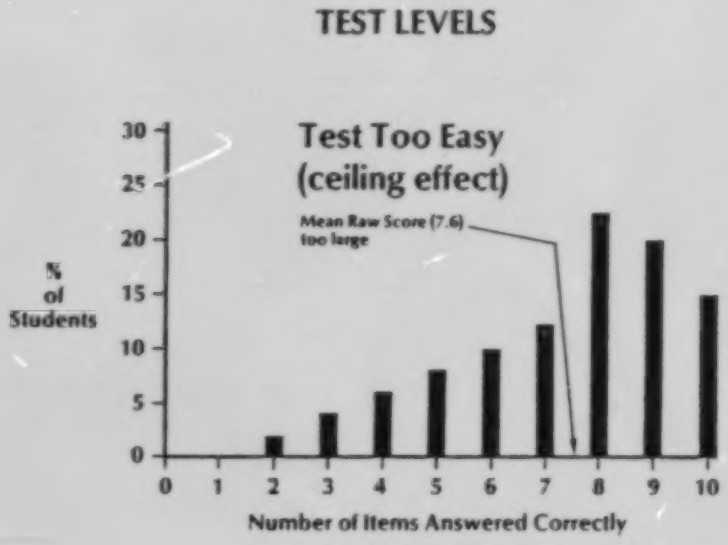


Figure 3. The ceiling effect which results from administering a test which is too easy

Evaluation Potential

If a large proportion of students are either "topping-out" or "bottoming-out" on the test either at pretest or posttest time, then the potential for demonstrating student gains between pretest and posttest when they actually occur is reduced. This happens because both ceiling and floor effects inaccurately estimate the actual level of student performance.

Ceiling effects depress the potential for exhibiting the upper level at which students are capable of functioning. For example, if the group responses from Figure 3 are used, a large proportion of the students (15%) are indicated as answering all ten items correct. If a longer test had been given with some more difficult items added, the chances are great that many of these students could have obtained an even higher score. Even those students with nine or fewer correct items could have increased their scores due to a combination of level of knowledge and chance responding. Therefore, we might be underestimating potential student performance at posttest time (Point D in Figure 4) when compared to the obtained posttest performance (Point C) and thereby reduce the potential for demonstrating actual student gains (difference between Point D and Point C when compared to Point A). In other words, because of a ceiling effect our obtained posttest (Point C) score will be at a maximum and still not assess the actual student performance level (Point D).

When ceiling effects are encountered at pretest time the effects on the evaluation outcomes can be even more serious. If the same test level is used again at posttest time the results could be either no growth or negative growth.

If a large proportion of students are "bottoming-out" on the test either at pretest or posttest time, then the potential for demonstrating actual student gains between pretest and posttest also is reduced. This occurs because a floor effect systematically overestimates the actual level of student performance.

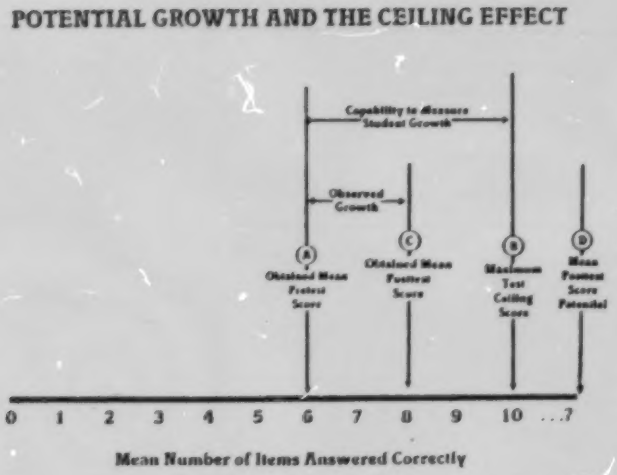


Figure 4. Reduced potential for student gains due to ceiling effect at posttest time

effect systematically overestimates the actual level of student performance. For example, if the group responses from Figure 2 are used, a large proportion of the students (55%) scored below the chance level score (assuming four-response, closed-ended items). If a longer test with some easier items added had been given, there is a high probability that many of these students would have missed even more items before arriving at their true score. Those students on the shorter test who received a real score which equated to the chance level would then have higher scores on the longer test when compared to those students who were functioning purely at chance level. Therefore, we might be overestimating student potential performance at pretest (Point A in Figure 5) when compared to the actual performance level (Point C) and reduce the potential for student gains (different between Point A and Point C when compared to Point B). In other words, because of the chance score and floor effects, our obtained pretest was higher than the actual student performance levels. This reduced our capability to measure the growth from pretest to posttest. If a floor effect is encountered at both pretest and posttest times, probably either no growth or

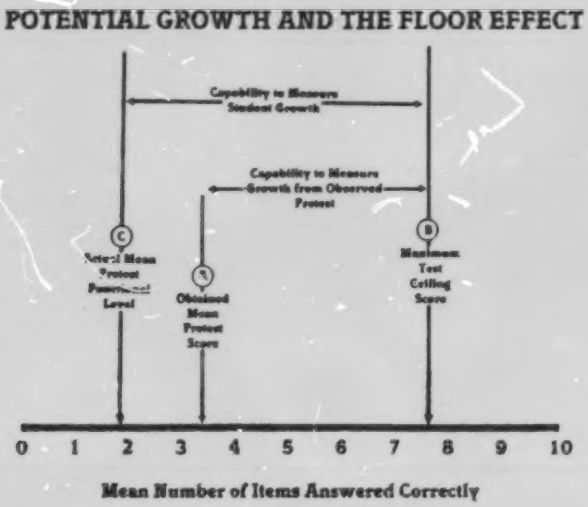

POTENTIAL GROWTH AND THE FLOOR EFFECT
Capability to Measure Student Growth
Capability to Measure Growth from Observed Pretest
C
Natural Mean Pretest Placement Level
O
Obtained Mean Pretest Score
B
Maximum Test Ceiling Score
0 1 2 3 4 5 6 7 8 9 10
Mean Number of Items Answered Correctly

Figure 5. Reduced potential for student gains due to floor effect at pretest time

negative growth will be demonstrated if the same test level is used for both administrations.

Test Reliability

The optimum test reliability is obtained when students correctly respond to slightly more than half of the test items (Stanley and Hopkins, 1972). Under these conditions there is no reason for administration of an out-of-level test as the in-level test is maximally measuring student achievement on this set of items. However, when a signficantly large number of students within a testing group correctly respond to considerably more or fewer items, the reliability of the test for this group of students diminishes. The test becomes more unreliable at the extreme ends of its score distribution (Nunnally, 1967). The relationship between reliability and percentage of correct items on a four-choice multiple choice test is roughly illustrated in Figure 6.

Thus, we should try to choose a test level on which the majority of the students will be receiving raw scores which correspond to approximately forty to seventy percent of the items correct.

OPTIMUM TEST RELIABILITY

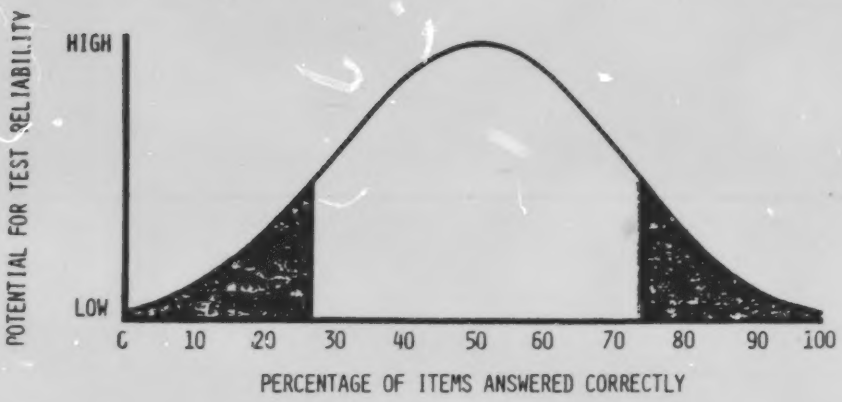


Figure 6. Relationship between percentage of correct items and test reliability

When Should A Student Be Tested At The Functional Level?

There are numerous instances when the on-level test may not be the best test for a given testing situation. In this section, five situations will be discussed where administration of an on-level test may not be the appropriate test level.

- When the On-Level Test is Much Too Easy or Difficult for the Student's Functional Level

There are many instances in which the on-level test may too difficult or too easy for a given student. In order to avoid floor or ceiling effects, it may be necessary to give a functional level test rather than the on-level test. A number of studies have indicated that functional level testing reduces the incidence of chance level scores (Ayrer and McNamara, 1973; Fuentes and Jeffress, n.d.). Barnes (n.d.) found that the incidence of chance scores

dropped from 25% of the population to 13% in one year and eventually to 9% in two years when switching to Functional Level Testing. Even more dramatic, Wick and Ward (1977) reported that chance scores dropped from 38% of the population to 9% in one year and to 8% in two years when switching to Functional Level Testing. However, Mendro (n.d.) found that out-of-level testing had not significantly reduced the percentage of chance level scores in his district.

Functional Level Testing has also been shown to reduce chance level scores with special populations. Cleland and Idstein (1980) found that more accurate test results could be obtained for special education students when they were tested two levels out-of-level, even though these results were still not in the desired range of scores. Similar results were obtained by Yoshida (1976) who found that out-of-level testing of EMRs reduced chance level score responding.

Therefore, if it is anticipated that a particular test level will be too easy or too difficult for a given student or group of students, that test level might produce results which are either at the chance level or at the ceiling level and not result in a valid assessment of achievement. Under these circumstances, an out-of-level test should be administered at the student's functional level(s).

- When the On-Level Test Will Provide Unreliable Results

As previously discussed, the potential for reliability of a test is highest when students are getting approximately forty to seventy percent of the items correct. Obviously, if a test is too easy or too difficult students will be achieving scores at the extreme upper or lower ends of the distribution which will reduce the possibility for obtaining a reliable result. Ayrer and McNamara found that "more reliable performance definitely resulted from the (Functional Level Testing) procedure..." (p. 79). This finding was dramatically verified by Wick (n.d.) who found that before Functional Level Testing students averaged

about 37-38 percent of the test items correct while with Functional Level Testing students averaged about 55-60 percent of the items correct.

Therefore, if it appears that administration of the on-level test will provide results which are unreliable, an out-of-level test which is more in keeping with the students' functional or instructional level should be considered.

- When the On-Level Test Does Not Match What the Student Is Being Taught

A fifth grader who is reading at the second grade level will receive very little special instruction with fifth grade content. Rather, if the program has been properly individualized this student should be receiving content at about his or her functional or instructional level. When this is happening, it is unfair to test the student with a test which measures growth in fifth grade content. Otherwise, the testing situation will have little content validity.

Numerous authors have identified a problem with content shifts in adjacent test levels (Ozenne, 1979; Raffeld, 1979). Therefore, the test level should be chosen so as to match the content which is being taught.

- When the On-Level Test Does Not Evaluate What the Program Is Designed to Teach

This consideration is similar to the one above except instead of being examined from the student learning point of view, it considers detecting growth in a group of students as a program effect. The city of Chicago had a tremendous problem in detecting Title I gains which they were sure were present. Upon examination of existing testing measures, it was discovered that about 65-80 percent of the Title I students were scoring at the chance level providing tremendous noise in the evaluation measures and reducing the potential for showing Title I effects. Indeed, when the switch was made to Functional Level Testing, the average drop in grade equivalents for Title I students by grade ranged

from seven months to 1.4 years. This switch to Functional Level Testing resulted in more accurate, albeit lower, pretest scores which allowed the gains which were occurring to be measured. This was primarily because the content being taught by the Title I program was at a much lower level than that which was being assessed by the on-level testing program (Wick, n.d.; Wick and Ward, 1977).

Of course, very few people would advocate evaluating growth in a math program with a reading test. However, by attempting to assess growth of fifth graders who are being instructed at the second grade level with fifth grade testing content will provide little opportunity for demonstrating the actual program effects. Therefore, the test levels chosen should match the actual instructional content which is being evaluated.

- When the On-Level Test Provides Little Useful Individual Diagnostic Information

When a test level is administered which is too easy or too difficult little useful information is available for diagnostic purposes about the capabilities of each student. For example, if on a forty item test a student is able to answer 39 of the items correctly, we know that this student performed very well on this test. However, little information has been obtained to guide further instruction for this student. It is not known whether other more difficult items are also known by this student. Likewise, a student who answers only 7 of 40 items correct will probably be performing at the chance level. Little information is provided by this test as to the actual skills and abilities that this student possesses. Under these circumstances, it is best that a test be administered which can provide useful diagnostic information for individual student program planning. Functional Level Testing is one way of obtaining this information.

The Controversy Over Functional Level Testing

The rosy picture painted above about the benefits of Functional Level Testing does not necessarily explain the whole view towards Functional Level Testing. There are a number of extremely vocal critics who caution against the use of Functional Level Testing. One of these critics has stated that Functional Level Testing procedures are "...faulty, inconsistent with the conclusions of current methodology, current measurement research, and should not be disseminated by AERA" (Jaeger, 1978a). He states further that he is "...convinced by work such as that of Long et al., and Slinde and Linn...that out-of-level testing and vertical equating generally lead to errors of such magnitude as to swamp achievement effects likely to be realized in Title I projects." (Jaeger, 1978b). Although another paper examines these issues in much more detail (Haenn, 1981), a brief review is in order.

The basic work referred to by Jaeger has been a finding that noncomparable results are obtained when using different levels of a test with the same students. However, in several of these studies (e.g., Pelavin and Barker, 1976; Slinde and Linn, 1977) either floor or ceiling effects occurred with the on- or off-level test negating the possibility for demonstrating equivalencies in scores from different levels. However, Slaughter and Gallas (1978) did find some inarticulation between levels on which no floor or ceiling effects occurred with either test level. Furthermore, examinations have demonstrated that the norms are least stable at the extreme points of the distribution (Ozenne, 1978; Applebaum, 1979). In addition, there is no consistency in the direction of the differences which result when different test levels are administered (Long, Schaffron and Kellogg, 1977; Ozenne, 1979). Finally, there are the content validity problems with articulating adjacent test levels that have been identified by Raffled (1979) and Ozenne (1979), but these content differences have

been rejected by Plake and Hoover (1979) as a possible cause for differences in grade equivalents obtained through different test levels.

On the positive side, "instructional level testing yielded a substantially more positive picture of reading improvement than did grade level testing" (Long, Schaffron and Kellogg, 1977, p. 208). Also, a number of studies have demonstrated a high correlation between scores obtained on the in-level and out-of-level tests (Pelavin and Barker, 1976; Slaughter and Gallas, 1978; Applebaum, 1979). In addition, two studies found little, if any, articulation problems between adjacent test levels (Barnes, n.d.; Wick and Ward, 1977). Wick and Ward (1977) attribute the similarity in scores between adjacent levels to the fact that they were using a test which has groups of items which overlap between each adjacent test level.

Therefore, there seems to be no concensus in opinion of the experts. It has been recommended that "out-of-level testing is a sound testing program which should be considered in any standardized testing program" (NWREL, 1978, p. 1). Further, Temp (1979) puts the issue in perspective when he states "...would you rather analyze data that <u>may</u> have some distortion due to the translation process in in-level percentile ranks or would you rather continue to analyze <u>chance</u> scores that are by definition merely random marks...?" (pp. 3-4). Long (1979), who is often cited by the opponents of Functional Level Testing, has stated that:

> ...on a purely logical basis it seems clear that off-level testing is usually at least testing something related to the area being tested whereas on-level testing often is seriously affected by floor effects as a large number of students who are at or below chance score levels. The basic issue, quite simply, is whether or not the scores obtained are comparable, and hence aggregatible (p. 7).

Thus, a decision needs to be made whether one wants to get more accurate information regarding the capabilities of the individual student and the effects of a

program or whether the results should be aggregated with a high degree of precision beyond that program.

Procedures for Conducting Functional Level Testing

When conducting Functional Level Testing the appropriate test levels to be administered must be determined as well as the method by which these various test levels actually will be administered. These two considerations often are interrelated since the determination of the appropriate level will be related to the administration procedures which are feasible. In the following sections the methods which can be utilized to determine appropriate testing levels are presented and the various procedures for administering Functional Level Testing are discussed. In addition, some guidelines are given as to how far one can safely test off-level without creating large articulation problems.

Determining Appropriate Test Levels for Students

There are several sources of information which can be used to select the appropriate test level for students. These sources of information are:

1. Previous test scores. If scores from the previous year or a previous test administration are available they should be examined carefully. If the mean of the scores for the student sample of interest indicates a floor effect was present, then the previous test was probably too difficult and a lower level of the current test should be selected. If the mean of the previous test scores suggests a ceiling effect, then the previous test was probably too easy for this group of students and a more difficult test level should be selected for the current testing. At the individual level, test scores should be examined to see if they are above the chance level but not approaching a test-ceiling level.

2. Current achievement level. If a reading or math achievement series is being utilized which provides approximate reading or math performance levels, these performance level indicators can be used as a rough estimate of the students' functional ability level.

3. Teacher judgement. Teachers or other specialists should be consulted and should review the content and difficulty of the various test levels to determine the appropriate test level for a given group of students or, in the case of individualized testing, for each student. The level(s) selected should be relevant to student needs, abilities, and stage of development. This is especially important in mathematics

where grade placement tends to be more rigid in relationship to test content. Barnes (n.d.) and Wick (n.d.) have reported that teachers can be effectively utilized in the test level selection process. However, Fuentes and Jeffress (n.d.) found that teachers generally underestimate student achievement levels.

4. Locator tests. Some test publishers provide short locator tests to aid the examiner in determining the appropriate test level for students. However, these locator tests should be used with caution as they also have been found to underestimate student achievement levels (Slaughter and Gallas, 1978).

5. Combinations of the above procedures. More accurate precision in the selection of appropriate test levels can be obtained by using a composite score based on a combination of two or more of the above sources of information.

Methods of Administering Functional Level Tests

In many instances an entire group of students (e.g., a group of Title I students) is very homogeneous but considerably different from their grade level peers. In these cases it may be appropriate to administer the same out-of-level test to the entire group of students. This is especially true if over three-quarters of the students are determined to need the same test level. This Functional Level Testing procedure is the easiest to administer with the fewest restrictions since the same level is being administered to all students. While a few students still may be taking a test level which is too difficult or too easy for them, the majority of students will receive an appropriate test level and the incidences of floor and ceiling effects will be reduced. For example, if 80% of the Title I students who are to be tested are functioning at one level below the recommended grade level test, this one-level-below test could be administered to all of the Title I students. Although the testing results for 20% of the population may have some problems, the overwhelming majority of the scores will be valid and reliable and the test results will provide a fairly accurate evaluation of the programmatic effects. If more accurate individual information is needed, those students who score at the extremes of the distribution can be retested later with a different test level.

Although administering a single out-of-level test to all students is the easiest procedure, it often may be impractical. For example, with a regular classroom containing average, accellerated, and disadvantaged students, at least three levels of the test may have to be administered: in-level, out-of-level one level above, and out-of-level one level below. Even within a group of disadvantaged students more than one level of the test may be appropriate depending upon the differences in abilities of the students and the grade range and content coverage designated for each test level.

With most tests the easiest way to accomplish administration of multiple test levels is by dividing the students into a number of groups with one group for each test level. Then, all students who will be taking the same level can be tested at the same time regardless of their grade placement or other factors.

A procedure called Individualized Testing is a special instance of multiple level testing. Several test publishers provide multiple test levels with identical instructions, practice items, and timing considerations for each level, even though the number of items may differ from one level to the next. In this way, multiple levels assigned on an individual basis can be administered at the same time to a widely disparate group of students. An alternative procedure is to test each student individually at his or her functional level.

How Far Should One Test Off-Level?

The decision of how far to test off-level will depend upon two factors: the width of each test level in terms of its nominal grade recommendation and content matter, and the norming procedures employed by the publishers. Some test publishers produce a different test level for every grade level. For other test publishers, test levels are designed to span several grade levels. In those instances when a test level is designed for a single grade, the publisher generally will have normative data for at least the adjacent grades on each side

and perhaps two grades out-of-level. In these instances, it is usually permissable to go up to two levels off level. One cautions against going further off level unless the publisher has used this practice during the norming period. For those publishers whose test levels span several grades, it is recommended that testing more than one test level off level should not occur unless the publisher collected information more than one level off level during the norming procedures.

Procedures for Converting Out-Of-Level Raw Scores to In-Level Norm Scores

One misconception regarding Functional Level Testing is that is gives students who are functioning below grade level an unfair advantage since they are administered an easier level of the test and will earn higher scores. However, this belief is invalid if the out-of-level score is interpreted in terms of the appropriate in-level reference or normative population. The major problem with Functional Level Testing to date has been the utilization of inappropriate reference populations (i.e., out-of-level norm groups rather than in-level norm groups). A student taking a lower and presumably easier level of the test may obtain a higher raw score than for the in-level test. However, the student must answer more items correctly on the out-of-level test than for the in-level test to obtain the same in-level norm-referenced score (e.g., grade equivalent, expanded standard score, or percentile rank).

Any test can be administered out of level if given to a population other than that for which it is nominally recommended. However, in order to appropriately interpret scores from administration of an out-of-level test a procedure must exist whereby the obtained out-of-level raw score can be compared with in-level norms. For example, a hypothetical reading test is designed so that third grade students take Level III, fourth grade students take Level IV, and fifth grade students take Level V. These are the nominally recommended, or

in-level, test levels. If a particular group of fifth graders was reading near the third grade level, the nominally recommended level (Level V) would probably be too difficult for these students and might result in chance level scores for a large proportion of the students. In order to avoid this possibility we might test these fifth grade students at their functional level by administering Level III. The obtained raw score would be an out-of-level raw score which would not be meaningful when compared with in-level raw scores obtained by the fifth grade students who took the nominally recommended level (Level V). It would also be meaningless to compare this out-of-level raw score to raw and normative scores for the third grade students since their raw and normative scores also were obtained through administration of an in-level (Level III) test. Therefore, a procedure must exist which allows an out-of-level raw score to be interpreted in-level (i.e., a method of taking the Level III out-of-level raw score in level to Level V so these fifth graders can be compared with other fifth graders).

There are three procedures available which allow out-of-level raw scores to be converted to in-level scores. The first type consists of specially designed tables by publishers which already have converted out-of-level raw scores to in-level normative scores. The other two types of procedures depend upon the existence of an expanded standard score scale which has been developed by the publisher to span all levels of the test. For example, the publisher of our hypothetical reading test has designed an expanded standard score scale to range from 200 to 800. An expanded standard score of 500 will be equal to different percentile equivalents for the different test levels as illustrated in Figure 7.

If the testing results are being machine-scored, then special care must be taken to ensure that the scoring service is aware that Functional Level Testing has occurred. This is usually accomplished through the use of the scoring service ordering blank. However, it is always a good idea to discuss your

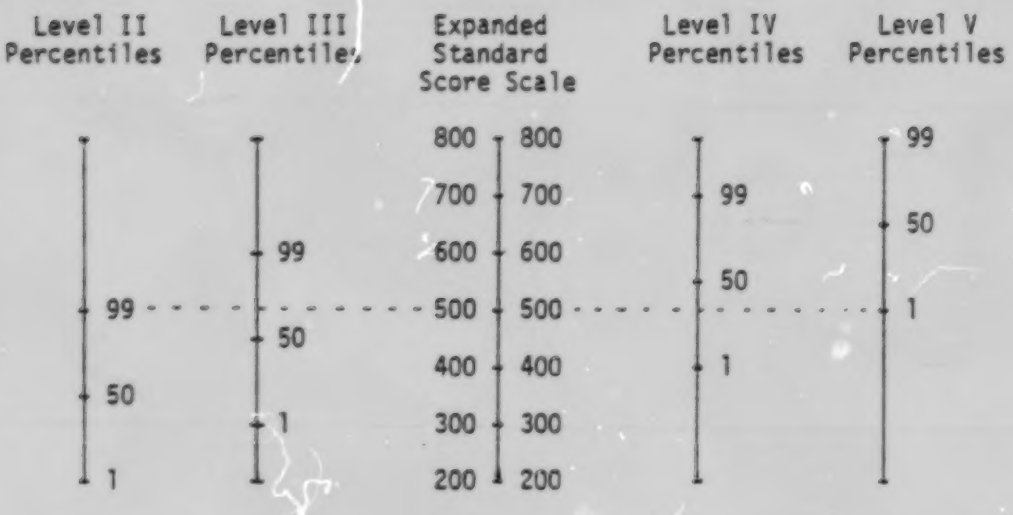


Figure 7. The concept of an expanded standard score scale illustrating the cross-test level equivalencies

Functional Level Testing program with your test representative before submitting the scores for machine-scoring. Once the machine-scored results have been returned to your district, a sample of these scores should be validated using the procedures described below so as to ensure that in-level conversions of out-of-level raw scores were completed. Only a small sample is necessary for this process since, if machine-scoring and score conversions were completed incorrectly, the errors can be readily detected.

If the testing results are to be hand-scored locally, then one of the following three methods must be used for converting the out-of-level raw scores to in-level norm scores.

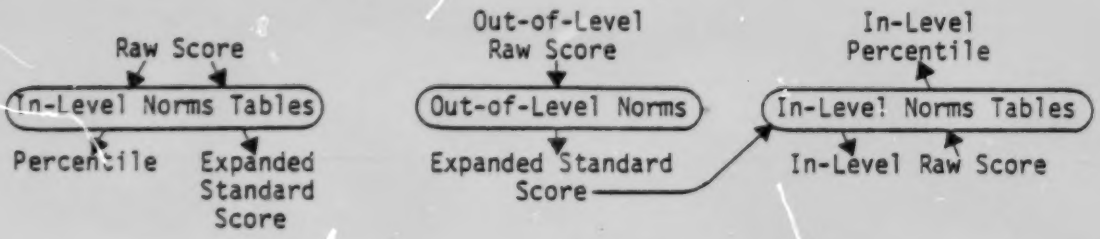
Out-Of-Level Norms Tables

Some publishers provide norms tables which contain out-of-level raw scores which have already been converted to in-level normative scores. When using these tables, the obtained out-of-level raw score is entered into these special

tables to find the tabled value, which will be a converted or in-level normative score.

Converting Out-Of-Level Raw Scores With Raw Score-To-Percentile Tables

This conversion method is used when publishers provide only raw score-to-percentile conversion tables. When testing in-level, the raw score can be directly converted to its percentile equivalent. (Figure 8a). However, when out-of-level testing is conducted the out-of-level raw score first must be



a. In-Level Score Conversion Procedures

b. Out-of-Level Score Conversion Procedures

Figure 8. Score conversion procedures for publishers which provide raw score-to-percentile conversion tables

converted to an expanded standard score using the out-of-level (actual test administration level) norms. Once the expanded standard score has been determined, it is entered into the in-level (nominally recommended level) norms, so that the in-level raw score corresponding to the out-of-level raw score and its expanded standard score equivalent can be determined. Theoretically, the in-level raw score is the raw score which would have been obtained had the in-level test been administered and no floor or ceiling effects encountered. Once the in-level raw score has been determined a simple table lookup using the in-level norms provides the in-level percentile corresponding to this converted in-level raw score. These steps are diagrammed in Figure 8b.

An example of the steps necessary to complete these conversions is provided in Figure 9 utilizing the Stanford Achievement Test. In this example, the pretest out-of-level raw score of 83 converts to an in-level raw score of 44.5 through the expanded standard score (called Scale Score) of 123. A sample worksheet for converting out-of-level raw scores with this conversion method is given in Appendix C.

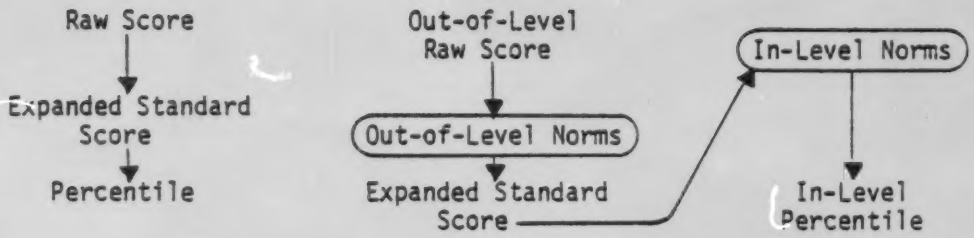
Converting Out-of-Level Raw Scores with Expanded Standard Score-to-Percentile Tables

This conversion method is used when publishers provide expanded standard score-to-percentile conversion tables. When testing in-level, the raw score must first be converted to an expanded standard score before it can be converted to a percentile (Figure 10a). When out-of-level testing is conducted the same three steps are used, but with two different sets of norms tables (Figure 10b). The out-of-level raw score is converted to an expanded standard score using the out-of-level (actual administration level) norms. Once the expanded standard score has been determined, it is entered directly into the in-level (nominally recommended level) norms to find the corresponding in-level percentile. Theoretically, the in-level percentile is the percentile equivalent which would have been obtained had the in-level test been administered and no floor or ceiling effects encountered. Since this conversion procedure is simpler when converting out-of-level scores to in-level normed scores, these tables should be used whenever a publisher provides both types of conversion tables.

An example of the steps necessary to complete these conversions is provided in Figure 11 utilizing the Metropolitan Achievement Tests as an example. A sample worksheet for converting out-of-level raw scores using this conversion method is given in Appendix D.

	Pretest	Posttest
DESCRIPTION		
Test Name:	Stanford Achievement Test	
Subtest Name:	Total Reading	
Expanded Standard Score Name:	Scale Score	
Student Placement:	Grade 4.2	Grade 4.7
Out-of-level instrument administered:	Primary II Level (appropriate for grades 2.7 and 3.2)	Primary II Level (appropriate for grades 2.7 and 3.2)
In-level instrument recommended by test publisher:	Primary III Level which is appropriate for grades 3.7 and 4.2.	Intermediate I Level, which is appropriate for grades 4.7 and 5.2.
PROCEDURES		
1.	Convert the out-of-level raw pretest scores to scale scores using Primary II Level tables. (As an example, use the out-of-level raw score of 83 which converts to the scale score of 123).	1. Convert the out-of-level raw posttest scores to scale scores using Primary II Level tables. (As an example, use the out-of-level raw score of 83 which converts to the scale score of 123).
2.	Using the Primary III Level tables for Beginning of Grade 4 which are recommended by the test publisher for grade 4.2, convert each pretest scale score to its corresponding in-level raw score. (e.g., 44.5)	2. Using the Intermediate I Level tables for End of Grade 4 which are recommended by the test publisher for grade 4.7, convert each posttest scale score to its corresponding in-level raw score. (e.g., 31.5)
3.	Using the Primary III Level raw-score-to-percentile table for Beginning of Grade 4 convert each pretest in-level raw score to its corresponding percentile. (e.g., 11.5)	3. Using the Intermediate I Level raw-scale-to-percentile table for End of Grade 4, convert each posttest in-level raw score to its corresponding percentile. (e.g., 5)

Figure 9. Example of the score conversion procedures for publishers which provide raw score-to-percentile conversion tables.



a. In-Level Score Conversion Procedures

b. Out-of-Level Score Conversion Procedures

Figure 10. Score conversion procedures for publishers which provide expanded standard score-to-percentile conversion tables.

	Pretest	Posttest
DESCRIPTION		
Test Name:	Metropolitan Achievement Tests	
Subtest Name:	Total Reading	
Expanded Standard Score Name:	Standard Score	
Student Placement:	Grade 4.2	Grade 4.7
Out-of-level instrument administered:	Primary II Level (appropriate for grades 2.7 and 3.1)	Primary II Level (appropriate for grades 2.7 and 3.1)
In-level instrument recommended by test publisher:	Elementary Level, which is appropriate for grades 3.7, 4.1 & 4.7.	Elementary Level, which is appropriate for grades 3.7, 4.1, and 4.7.
PROCEDURES		
	1. Convert the out-of-level raw pretest scores to standard scores using Primary II Level tables. (As an example, use the out-of-level raw score of 55 which converts to the standard score of 52).	1. Convert the out-of-level raw posttest scores to standard scores using Primary II Level tables. (As an example, use the out-of-level raw score of 73 which converts to the standard score of 59).
	2. Using the Elementary Level standard score-to-percentile table for Beginning of Grade 4 convert each pretest in-level standard score to its corresponding percentile. (e.g., 16)	2. Using the Elementary Level standard score-to-percentile table for End of Grade 4, convert each posttest in-level standard score to its corresponding percentile. (e.g., 20)

Figure 11. Example of the score conversion procedures for publishers which provide expanded standard score-to-percentile conversion tables.

REFERENCES

Applebaum, W.R. "Effects of Off-Level Assignment on ITBS Grade Equivalent Scores." Paper presented at AERA Annual Meetings, Los Angeles, CA, April, 1979.

Ayrer, J.E. and T.C. McNamera. "Survey Testing on an Out-of-Level Basis." Journal of Educational Measurement, 1973, 10(2), pp.79-83.

Barnes, J. "A Report on Individualized Testing in Atlanta." Undated paper.

Burton, R. "Chance-Level Scores, Response Models and the Mann-Whitney U Test." Paper presented at the Eastern Educational Research Association Annual Meeting, February, 1979. Charleston, SC.

Butera, B.J. and P.C. Raffeld. "Investigation of Vertical Scaling for Out-of-Level Testing." Paper presented at AERA Annual Meeting, San Francisco, CA, April 1979.

Cleland, W.E. and P.M. Idstein. "In-Level Versus Out-of-Level Testing of Sixth Grade Special Education Students." Paper presented at NCME Annual Meeting, Boston, MA, April 1980.

Doherty, W. "Restandardization Study." The Combined ESAA Pilot and Basic Program Evaluations:Final Study Design (Volume III). Santa Monica, CA: Systems Development Corporation, February 1977. As cited in Ozenne (1978).

Fisher, J.A. "The Use of Out-of-Grade Tests with Retarded and Accelerated Readers." Unpublished Doctoral Dissertation, State University of Iowa, 1961.

Fuentes, E.J. and E.L. Jeffress. "The Use of Teacher Estimates of Children's Achievement in Selecting Appropriate Test Levels vs. Blanket Out-of-Level Testing: The Problem of an Efficient Method of Matching the Students Functional Level with the Proper Test Level." Research Triangle Park, NC: Research Triangle Institute, Center for Educational Research and Evaluation. Undated paper.

Gulliksen, H. Theory of Mental Tests. New York: Wiley and Sons, 1950.

Haenn, J.F. "Functional Level Testing: Is It Supported or Refuted by the Literature." Paper in press in development, 1981.

Haenn, J.F., and D.C. Proctor. "A Practitioner's Guide to Out-of-Level Testing." Paper presented at AERA Annual Meeting, Toronto, Canada, March 1978.

Jaeger, R.M. Personal correspondence. October 31, 1978a.

Jaeger, R.M. (1978b). Personal correspondence. November 3, 1978b.

Long, J.V. "Out-of-Level Survey Testing in Title I -- Further Empirical Examination." Paper presented at AERA Annual Meeting, San Francisco, CA, April 1979.

Long, J.V., J.A. Schaffran and T.M. Kellogg. "Effects of Out-of-Level Survey Testing on Reading Achievement Scores of Title I, ESEA Students." Journal of Educational Measurement, 1977, 14(3), pp.203-213.

Mendro, R. "A Report on Individualized Testing in Dallas." Undated paper.

Nunnally, J.C. Psychometric Theory. New York: McGraw-Hill, 1967.

NWREL. "Out-of-Level Testing: Rationale to Implementation." Portland, OR: Northwest Regional Educational Laboratory, Title I Evaluation Technical Assistance Center, 1978.

Ozenne, D. "You May Not Get What You Think You Are Getting: What Test Scores Mean in Out-of-Level Testing." Paper presented at AERA Annual Meeting, San Francisco, CA, April 1979.

Ozenne, D. "The Effect of Vertical Scaling Imprecision in the Estimation of Title I Project Gains." Paper presented at AERA Annual Meeting, Toronto, Canada, March 1978.

Pelavin, S.H. and P. Barker. "A Study of the Generalizability of the Results of a Standardized Achievement Test." Paper presented at AREA Annual Meeting, New York City, 1976.

Plake, B.S. and H.D. Hoover. "The Comparability of Equal Raw Scores Obtained from In-Level and Out-of-Level Testing: One Source of the Discrepancy Between In-Level and Out-of-Level Grade Equivalent Scores." Journal of Educational Measurement, 1979, 16(4), pp.271-278.

Raffeld, P. "Functional Level Testing and the Content Shift." Paper presented at AERA Annual Meeting, San Francisco, CA., April 1979.

Roberts. A.O.H. Out-of-Level Testing. Mountain View, CA: RMC Research Corporation, January 1978a.

Roberts, S.J. Test Floor and Ceiling Effects. Mountain View, CA: RMC Research Corporation, January 1978b.

Silverman, B. "Test Bias and Ability Level Testing." Journal of School Psychology, 1979, 17(3), pp.255-259.

Slaughter, H.B. and E.J. Gallas. "Will Out-of-Level Norm-Referenced Testing Improve the Selection of Program Participants and the Diagnosis of Reading Comprehension in ESEA Title I Programs?" Paper presented at AERA Annual Meeting, Toronto, Canada, March 1978.

Slinde, J.A. and R.L. Linn. "Vertically Equated Tests: Fact or Phantom?" Journal of Educational Measurement, 1977, 14(1), pp. 23-32.

Stamman, D. "Chance Score Identification." Austin, TX: Recourse Development Institute. Undated draft copy.

Stanley, J.C. and K.D. Hopkins. Education and Psychological Measurement and Evaluation. Englewood Cliffs, NJ: Prentice-Hall, Inc., 1972.

Temp, G.E. "Out-of-Level Testing Works." Paper presented at AERA Annual Meeting, San Francisco, CA, April 1979.

USED. Policy Manual. Chapter 10--Evaluation. Washington, DC: U.S. Department of Education, December 1980.

Wick, J. "A Report on Individualized Testing in Chicago." Undated paper.

Wick, J.W. and F. Ward. "Testing Students At Functioning Reading Level: A Two-Year Report From Chicago." (1977)

Yoshida, R.K. "Out-of-Level Testing of Special Education Students with a Standardized Achievement Battery." Journal of Educational Measurement, 1976, 13(3), pp.215-221.

APPENDIX A

Chance Score Grade Equivalents and Percentiles of Various Standardized Tests

This appendix contains the chance raw scores and their corresponding grade equivalent scores and percentile ranks for reading, mathematics, and language arts subtests of tests which are commonly used in program evaluation for grades 2-6. Indicated for the subtests of each level of each test appropriate for this grade span are the total number of items, the number of alternatives per item, the chance raw score and the grade equivalent corresponding to this chance raw score. The percentile equivalents of this chance score for the lowest and highest in-level tabled values available for grades between the beginning of grade 2 (B-2) and the end of grade 6 (E-6) are also provided. Where the number of alternatives varied among items within a subtest, the number of items is indicated for each number of alternatives. In some instances, a level was designed to go below or above the indicated tabled values for percentiles. The nominally recommended administration levels for each test and level are given on the following page. These nominally recommended levels should especially be considered carefully when examining tests which have multilevel editions, such as the Iowa Tests of Basic Skills and the SRA Achievement Series.

Nominally Recommended Grade Levels for Test Levels of Ten Achievement Tests Widely Used in Grades 2-6

Test Name	Level Name	Nominally Recommended Grades Lowest Grade	Highest Grade
California Achievement Tests (1977)	12	1.5*	2.9
	13	2.5	3.9
	14	3.5	4.9
	15	4.5	5.9
	16	5.5	6.9
Comprehensive Tests of Basic Skills (1973)	C	1.6	2.9
	1	2.5	4.9
	2	4.5	6.9
Gates-MacGinitie Reading Tests (1978)	B	B-2**	E-2***
	C	B-3	E-3
	D	B-4	E-6
Iowa Tests of Basic Skills (1978)	Primary 7	1.7	2.6
	Primary 8	2.7	3.5
	9	B-3	E-3
	10	B-4	E-4
	11	B-5	E-5
	12	B-6	E-6
Metropolitan Achievement Tests (1978)	Primary I	1.5	2.4
	Primary II	2.5	3.4
	Elementary	3.5	4.9
	Intermediate	5.0	6.9
SRA Achievement Series (1978)	B	E-1	B-3
	C	E-2	B-4
	D	E-3	B-5
	E	E-4	B-6
	F	E-5	B-8
Stanford Achievement Test (1973)	Primary I	1.5	2.4
	Primary II	2.5	3.4
	Primary III	3.5	4.4
	Intermediate I	4.5	5.4
	Intermediate II	5.5	6.9
Stanford Diagnostic Reading Test (1976)	Red	1.6	3.5
	Green	2.6	5.5
	Brown	4.6	9.5
Stanford Diagnostic Mathematics Test (1976)	Red	1.6	4.5
	Green	3.6	6.5
	Brown	4.6	9.5
Sequential Tests of Educational Progress III (1979)	Intermediate	E-3	B-4
	Intermediate F	E-4	B-5
	Intermediate G	E-5	B-6

* First digit indicates grade; second digit indicates month within grade where September is month 0.
** B = beginning of grade.
*** E = end of grade.

Sequential Tests of Educational Progress III (1979)	Circus Primary C	1.5	2.5
	Circus Primary D	2.5	3.5
	Intermediate E	3.5	4.5
	Intermediate F	4.5	5.5
	Intermediate G	5.5	6.5
	Intermediate H	6.5	7.5

California Achievement Tests, Level 12 (Form C, 1977)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Phonic Analysis	25	10/3 15/4	7	-	B-2	10	B-3	4
Structural Analysis	11	4	3	-	B-2	11	B-3	2
Reading Vocabulary	15	4	4	1.4	B-2	18	B-3	7
Reading Comprehension	20	4	5	1.6	B-2	21	B-3	7
Total Reading	71	10/3 61/4	19	1.1	B-2	8	B-3	1
Language Mechanics	20	4	5	1.2	B-2	9	B-3	2
Language Expression	20	2	10	1.5	B-2	19	B-3	7
Total Language	40	20/2 20/4	15	1.1	B-2	9	B-3	2
Math Computation	26	4	7	1.6	B-2	23	B-3	3
Math Concepts and Applications	40	4	10	0.7	B-2	3	B-3	1
Total Mathematics	66	4	17	1.1	B-2	5	B-3	1

California Achievement Tests, Level 13 (Form C, 1977)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Phonic Analysis	20	5/3 15/4	5	-	E-2	9	B-4	-
Structural Analysis	11	4	3	-	E-2	10	B-4	-
Reading Vocabulary	15	5/3 10/4	4	1.6	E-2	13	B-4	3
Reading Comprehension	27	7/3 20/4	7	1.5	E-2	35	B-4	2
Total Reading	73	17/3 56/4	20	1.5	E-2	8	B-4	2
Math Computation	40	5	8	2.3	E-2	25	B-4	2
Math Concepts and Applications	45	5	9	1.6	E-2	11	B-4	3
Total Mathematics	85	5	17	1.7	E-2	11	B-4	2
Language Mechanics	20	10/4 10/5	5	1.8	E-2	16	B-4	4
Language Expression	26	11/2 11/3 4/4	10	1.7	E-2	14	B-4	3
Total Language	46	11/2 11/3 14/4 10/5	15	1.7	E-2	13	B-4	2

California Achievement Tests, Level 14 (Form C, 1977)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Reading Vocabulary	30	5/3 25/4	8	2.6	E-3	25	B-5	6
Reading Comprehension	40	4	10	2.5	E-3	22	B-5	7
Total Reading	70	5/3 65/4	18	2.6	E-3	22	B-5	5
Language Mechanics	25	5	5	2.1	E-3	12	B-5	4
Language Expression	38	8/2 12/3 18/4	13	2.3	E-3	18	B-5	7
Total Language	63	8/2 12/3 18/4 25/5	18	2.3	E-3	13	B-5	4
Math Computation	40	5	8	3.0	E-3	15	B-5	2
Math Concepts and Applications	45	5	9	2.4	E-3	14	B-5	3
Total Mathematics	85	5	17	2.5	E-3	11	B-5	2

California Achievement Tests, Level 15 (Form C, 1977)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Reading Vocabulary	30	5/3 25/4	8	2.9	E-4	12	B-6	5
Reading Comprehension	40	4	10	2.8	E-4	16	B-6	6
Total Reading	70	5/3 65/4	18	2.8	E-4	12	B-6	4
Language Mechanics	25	5	5	2.4	E-4	8	B-6	4
Language Expression	38	8/2 12/3 18/4	13	2.3	E-4	9	B-6	4
Total Language	63	8/2 12/3 18/4 25/5	18	2.4	E-4	6	B-6	3
Math Computation	40	5	8	3.5	E-4	11	B-6	2
Math Concepts and Applications	45	5	9	3.0	E-4	12	B-6	4
Total Mathematics	85	5	17	3.2	E-4	8	B-6	2

California Achievement Tests, Level 16 (Form C, 1977)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Reading Vocabulary	30	5/3 25/4	8	3.4	E-5	11	E-6	6
Reading Comprehension	40	4	10	3.4	E-5	13	E-6	8
Total Reading	70	5/3 65/4	18	3.3	E-5	10	E-6	5
Language Mechanics	25	5	5	2.3	E-5	4	E-6	4
Language Expression	38	8/2 12/3 18/4	13	2.5	E-5	7	E-6	6
Total Language	63	8/2 12/3 18/4 25/5	18	2.5	E-5	4	E-6	3
Math Computation	40	5	8	3.9	E-5	9	E-6	4
Math Concepts and Applications	45	5	9	3.2	E-5	8	E-6	4
Total Mathematics	85	5	17	3.5	E-5	6	E-6	3

Comprehensive Test of Basic Skills, Level C (Form S, 1973)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Vocabulary	33	4	8	--	B-2	6	E-2	2
Sentences	23	4	6	--	B-2	17	E-2	8
Passages	18	4	5	--	B-2	24	E-2	9
Total Reading	74	4	19	1.3	B-2	6	E-2	3
Expression	22	4	6	--	B-2	15	E-2	7
Spelling	34	2	17	--	B-2	19	E-2	9
Mechanics	23	4	6	--	B-2	23	E-2	12
Total Language	79	34/2 45/4	28	0.7	B-2	13	E-2	4
Math Computation	28	4	7	--	B-2	14	E-2	5
Math Concepts/Applic.	25	4	6	--	B-2	5	E-2	2
Total Mathematics	53	4	13	0.9	B-2	6	E-2	2
Comprehensive Test of Basic Skills, Level 1 (Form S, 1973)								
Vocabulary	40	4	10	1.9	B-2	42	E-4	4
Comprehension	45	4	11	2.1	B-2	50	E-4	6
Total Reading	85	4	21	1.9	B-2	39	E-4	4
Spelling	44	2	22	2.1	B-2	49	E-4	8
Mechanics	20	4	5	1.7	B-2	35	E-4	5
Expression	30	4	8	1.9	B-2	39	E-4	5
Total Language	94	44/2 50/4	35	2.0	B-2	44	E-4	4
Math Computation	48	4	12	2.2	B-2	57	E-4	2
Math Concepts	25	4	6	1.6	B-2	32	E-4	3
Math Applications	25	4	6	2.2	B-2	53	E-4	6
Total Mathematics	98	4	25	1.9	B-2	39	E-4	2

Comprehensive Test of Basic Skills, Level 2 (Form S, 1973)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Vocabulary	40	4	10	2.5	B-4	16	E-6	4
Comprehension	45	4	11	2.6	B-4	16	E-6	6
Total Reading	85	4	21	2.3	B-4	12	E-6	3
Spelling	50	2	25	2.6	B-4	19	E-6	8
Mechanics	20	4	5	2.0	B-4	11	E-6	5
Expression	35	4	9	2.1	B-4	10	E-6	5
Total Language	105	50/2 55/4	39	2.3	B-4	12	E-6	3
Math Computation	48	4	12	3.0	B-4	15	E-6	3
Math Concepts	25	4	6	2.0	B-4	9	E-6	3
Math Applications	25	4	6	2.9	B-4	22	E-6	6
Total Mathematics	98	4	25	2.3	B-4	9	E-6	2

Gates-MacGinitie Reading Test, Level B (Form 1, 1978)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Vocabulary	45	4	11	1.4	B-2	14	E-2	3
Comprehension	40	4	10	1.4	B-2	11	E-2	3
Gates-MacGinitie Reading Test, Level C (Form 1, 1978)	**Number of Items**	**Number of Alternatives**	**Chance Raw Score**	**Grade Equiv.**	**Grade**	**%ile**	**Grade**	**%ile**
Vocabulary	45	4	11	1.9	B-3	15	E-3	7
Comprehension	44	4	11	1.6	B-3	8	E-3	3
Gates-MacGinitie Reading Test, Level D (Form 1, 1978)	**Number of Items**	**Number of Alternatives**	**Chance Raw Score**	**Grade Equiv.**	**Grade**	**%ile**	**Grade**	**%ile**
Vocabulary	45	5	9	2.5	B-4	11	E-6	1
Comprehension	43	4	1	2.4	B-4	14	E-6	3

Iowa Test of Basic Skills, Primary Level 7 (Form 7, 1978)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Listening	32	3	11	K.5	B-2	2	E-2	1⁻
Word Analysis	49	3	16	K.3	B-2	1	E-2	1⁻
Vocabulary	30	5/3 25/4	8	K.9	B-2	10	E-2	3
Reading	66	32/2 23/3 11/4	27	1.2	B-2	14	E-2	4
Language								
Spelling	27	3	9	1.1	B-2	11	E-2	6
Capitalization	66	*	--	---	---	--	---	-
Punctuation	50	*	--	---	---	--	---	-
Usage	24	3	8	K.7	B-2	8	E-2	2
Total Language	167	*	--	---	---	--	---	-
Mathematics								
Math Concepts	33	3	11	K.6	B-2	1⁻	E-2	1⁻
Math Problem-Solving	22	4	6	K.7	B-2	3	E-2	1⁻
Math Computation	26	4	7	1.0	B-2	1	E-2	1⁻
Total Mathematics	81	33/3 48/4	23	K.8	B-2	1⁻	E-2	1⁻

Iowa Test of Basic Skills, Primary Level 8 (Form 7, 1978)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Listening	32	3	11	1.2	E-2	6	E-3	1⁻
Word Analysis	57	3	19	1.0	E-2	6	E-3	1⁻
Vocabulary	30	4/3 26/4	8	1.5	E-2	14	E-3	3
Reading	67	16/2 28/3 23/4	23	1.5	E-2	9	E-3	2
Language								
Spelling	29	3	10	1.2	E-2	7	E-3	2
Capitalization	75	*	--	---	---	--	---	-
Punctuation	68	*	--	---	---	--	---	-
Usage	23	3	8	1.3	E-2	13	E-3	3
Total Language	195	*	--	---	---	--	---	-
Mathematics								
Math Concepts	36	2/3 34/4	9	1.4	E-2	3	E-3	1⁻
Math Problem-Solving	24	4	6	1.1	E-2	4	E-3	1⁻
Math Computation	28	4	7	1.6	E-2	2	E-3	1⁻
Total Mathematics	88	2/3 86/4	22	1.4	E-2	1	E-3	1⁻

* Number of responses varies by item.

Iowa Test of Basic Skills, Level 9 (Form 7, 1978)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Vocabulary	30	4	8	2.0	B-3	18	E-3	10
Reading Comprehension	44	4	11	2.0	B-3	14	E-3	6
Language								
Spelling	30	5	6	1.8	B-3	12	E-3	7
Capitalization	28	4	7	2.2	B-3	20	E-3	10
Punctuation	28	4	7	2.4	B-3	26	E-3	14
Usage	27	4	7	2.1	B-3	25	E-3	15
Total Language	113	83/4 30/5	27	2.1	B-3	14	E-3	6
Mathematics								
Math Concepts	28	4	7	2.1	B-3	11	E-3	4
Math Problem-Solving	23	4	6	2.2	B-3	20	E-3	10
Math Computation	39	4	10	2.6	B-3	16	E-3	5
Total Mathematics	90	4	23	2.3	B-3	11	E-3	3

Iowa Test of Basic Skills, Level 10 (Form 7, 1978)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Vocabulary	36	4	9	2.5	B-4	13	E-4	7
Comprehension	49	4	12	2.7	B-4	14	E-4	7
Language								
Spelling	36	5	7	2.1	B-4	8	E-4	5
Capitalization	29	4	7	2.5	B-4	11	E-4	6
Punctuation	29	4	7	2.7	B-4	15	E-4	9
Usage	29	4	7	2.4	B-4	15	E-4	9
Total Language	123	87/4 36/5	29	2.4	B-4	7	E-4	3
Mathematics								
Math Concepts	32	4	8	2.9	B-4	12	E-4	5
Math Problem-Solving	25	4	6	2.8	B-4	15	E-4	8
Math Computation	42	4	11	3.2	B-4	10	E-4	4
Total Mathematics	99	4	25	3.0	B-4	10	E-4	3

Iowa Test of Basic Skills, Level 11 (Form 7, 1978)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Vocabulary	39	4	10	3.5	B-5	16	E-5	10
Comprehension	54	4	14	3.6	B-5	17	E-5	10
Language								
Spelling	40	5	8	2.7	B-5	8	E-5	5
Capitalization	30	4	8	3.1	B-5	10	E-5	6
Punctuation	30	4	8	3.6	B-5	19	E-5	13
Usage	30	4	8	3.2	B-5	17	E-5	11
Total Language	130	90/4 40/5	31	3.2	B-5	9	E-5	5
Mathematics								
Math Concepts	37	4	9	3.7	B-5	12	E-5	6
Math Problem-Solving	27	4	7	3.7	B-5	16	E-5	10
Math Computation	45	4	11	3.8	B-5	9	E-5	4
Total Mathematics	109	4	27	3.7	B-5	8	E-5	3

Iowa Test of Basic Skills, Level 12 (Form 7, 1978)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Vocabulary	42	4	11	4.2	B-6	14	E-6	9
Comprehension	56	4	14	4.1	B-6	13	E-6	8
Language								
Spelling	42	5	8	3.1	B-6	6	E-6	4
Capitalization	30	4	8	4.0	B-6	13	E-6	10
Punctuation	30	4	8	3.9	B-6	14	E-6	9
Usage	30	4	8	3.6	B-6	13	E-6	9
Total Language	132	90/4 42/5	31	3.7	B-6	8	E-6	4
Mathematics								
Math Concepts	40	4	10	4.4	B-6	11	E-6	6
Math Problem-Solving	29	4	7	4.6	B-6	18	E-6	12
Math Computation	45	4	11	4.3	B-6	6	E-6	3
Total Mathematics	114	4	29	4.4	B-6	8	E-6	3

Metropolitan Achievement Test, Primary 1 (Form JS, 1978)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Reading Comprehension	55	3	18	1.3	B-2	11		
Mathematics	40	4	10	PK*	B-2	1		
Language	40	2/2 17/3 21/4	12	K.0	B-2	1		

Metropolitan Achievement Test, Primary II (Form JS, 1978)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Reading Comprehension	55	4	14	1.7	E-2	9	B-3	5
Mathematics	45	4	11	1.0	E-2	2	B-3	1
Language	55	22/3 33/4	16	1.2	E-2	2	B-3	1

* PreKindergarten

Metropolitan Achievement Test, Elementary (Form JS, 1978)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Reading Comprehension	60	4	15	1.9	E-3	5	E-4	1
Mathematics	50	1/3 49/4	13	1.9	E-3	4	E-4	1
Language	60	24/3 36/4	17	1.1	E-3	1	E-4	1

Metropolitan Achievement Test, Intermediate (Form JS, 1978)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Reading Comprehension	60	4	15	2.4	B-5	8	E-6	4
Mathematics	50	4	13	2.9	B-5	10	E-6	4
Language	60	4	15	2.0	B-5	3	E-6	1

SRA Achievement Series, Level B (Form 1, 1978)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Auditory Discrimination	25	2	13	0.1	B-2	10	B-3	7
Letters and Sounds	25	4	6	0.1	B-2	1	B-3	1
Listening Comprehension	20	4	5	0.1	B-2	1	B-3	1
Vocabulary	25	4	6	0.1	B-2	1	B-3	1
Reading Comprehension	24	4	6	0.1	B-2	7	B-3	2
Total Reading	119	25/2 94/4	36	0.1	B-2	1	B-3	1
Math Concepts	30	4	8	0.1	B-2	1	B-3	1
Math Computation	27	4	7	0.1	B-2	2	B-3	1
Total Mathematics	57	4	14	0.1	B-2	1	B-3	1

SRA Achievement Series, Level C (Form 1, 1978)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Auditory Discrimination	00	0	00	0.0	000	0	000	0
Letters and Sounds	20	4	5	0.8	E-2	2	B-4	1
Listening Comprehension	20	4	5	0.1	E-2	1	B-4	1
Vocabulary	25	4	6	0.1	E-2	4	B-4	2
Reading Comprehension	24	4	6	0.1	E-2	2	B-4	1
Total Reading	89	4	22	0.3	E-2	1	B-4	1
Math Concepts	35	3/3 32/4	9	0.3	E-2	2	B-4	1
Math Computation	27	4	7	0.1	E-2	3	B-4	1
Total Mathematics	62	3/3 59/4	16	0.6	E-2	1	B-4	1
Language Mechanics	20	4	5	0.1	E-2	4	B-4	1
Language Usage	20	4	5	0.1	E-2	5	B-4	2
Spelling	29	4	7	0.1	E-2	5	B-4	1
Total Language Arts	69	4	17	0.1	E-2	2	B-4	1

SRA Achievement Series, Level D (Form 1, 1978)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Reading Vocabulary	30	4	8	1.8	E-3	7	B-5	6
Reading Comprehension	26	4	7	1.8	E-3	8	B-5	5
Total Reading	56	4	14	1.3	E-3	4	B-5	2
Math Concepts	35	4	9	1.6	E-3	4	B-5	3
Math Computation	35	4	9	2.3	E-3	5	B-5	3
Total Mathematics	70	4	18	2.0	E-3	3	B-5	2
Language mechanics	30	4	8	0.1	E-3	4	B-5	2
Language usage	30	4	8	0.3	E-3	7	B-5	5
Spelling	30	4	8	0.9	E-3	7	B-5	5
Total Language Arts	90	4	23	0.1	E-3	3	B-5	2

SRA Achievement Series, Level E (Form 1, 1978)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Reading Vocabulary	40	4	10	1.6	E-4	6	B-6	4
Reading Comprehension	50	4	13	1.8	E-4	6	B-6	4
Total Reading	90	4	23	1.5	E-4	5	B-6	3
Math Concepts	30	4	8	1.6	E-4	4	B-6	2
Math Computation	40	5	8	2.4	E-4	4	B-6	1
Math Problem-Solving	30	4	8	2.1	E-4	11	B-6	7
Total Mathematics	100	60/4 40/5	23	2.1	E-4	3	B-6	1
Language Mechanics	30	4	8	0.1	E-4	4	B-6	3
Language Usage	40	4	10	0.1	E-4	4	B-6	4
Spelling	30	4	8	2.0	E-4	11	B-6	8
Total Language Arts	100	4	25	0.1	E-4	2	B-6	1

SRA Achievement Series, Level F (Form 1, 1978)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Reading Vocabulary	40	4	10	2.7	E-5	13	E-6	7
Reading Comprehension	50	4	13	2.3	E-5	10	E-6	6
Total Reading	90	4	23	2.1	E-5	10	E-6	5
Math Concepts	30	4	8	2.7	E-5	12	E-6	5
Math Computation	40	5	8	3.5	E-5	9	E-6	5
Math Problem-Solving	30	4	8	2.7	E-5	19	E-6	11
Total Mathematics	100	60/4 40/5	23	3.0	E-5	7	E-6	2
Language Mechanics	30	4	8	0.9	E-5	7	E-6	5
Language Usage	40	4	10	0.5	E-5	5	E-6	3
Spelling	30	4	8	2.9	E-5	14	E-6	10
Total Language Arts	100	4	25	1.1	E-5	5	E-6	3

Stanford Achievement Test, Primary Level I (Form A, 1973)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Vocabularly	37	3	12	K.1	B-2	6		
Reading (Part A)	45	3	15	1.0	B-2	8		
Reading (Part B)	42	3	14	1.2	B-2	11		
Word Study Skills	60	3	20	K.9	B-2	2		
Total Reading	147	3	49	K.8	B-2	4		
Math Concepts	32	4	8	K.1	B-2	1		
Math Computation	32	4	8	K.6	B-2	4		
Total Mathematics	64	4	16	K.8	B-2	1		
Stanford Achievement Test, Primary Level III (Form A, 1973)								
Vocabularly	45	4	11	1.0	E-3	4	B-4	2
Reading Comprehension	70	4	18	1.7	E-3	4	B-4	4
Word Study Skills	55	31/3 24/4	16	1.8	E-3	6	B-4	10
Total Reading	125	31/3 94/4	34	1.6	E-3	4	B-4	4
Math Concepts	32	4	8	2.1	E-3	8	B-4	8
Math Computation	36	24/3 12/5	10	2.2	E-3	8	B-4	8
Math Applications	28	5	6	2.1	E-3	8	B-4	8
Total Mathematics	96	24/3 32/4 40/5	24	2.0	E-3	4	B-4	4

Stanford Achievement Test, Primary Level II (Form A, 1973)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Vocabulary	37	3	12	1.0⁻	E-2	4	B-3	2
Reading (Part A)	45	3	15	1.4	E-2	4	B-3	4
Reading (Part B)	48	4	12	1.6	E-2	6	B-3	6
Word Study Skills	65	3	22	1.3	E-2	4	B-3	1
Total Reading	158	110/3 48/4	49	1.2	E-2	2	B-3	2
Math Concepts	35	1/3 31/4 1/5 2/23	8	1.0⁻	E-2	2	B-3	1
Math Computation	37	8/3 29/5	8	1.0⁻	E-2	2	B-3	2
Math Applications	28	11/4 17/5	6	1.3	E-2	4	B-3	2
Total Mathematics	100	9/3 42/4 47/5 2/23	23	1.0	E-2	1	B-3	1

Stanford Achievement Test, Intermediate Level I (Form A, 1973)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Vocabularly	50	4	13	2.2	E-4	6	B-5	4
Reading Comprehension	72	4	18	2.7	E-4	11	B-5	10
Word Study Skills	55	25/3 30/4	16	2.1	E-4	10	B-5	10
Total Reading	127	25/3 102/4	34	2.5	E-4	8	B-5	6
Math Concepts	32	4	8	3.0	E-4	12	B-5	10
Math Computation	40	18/3 22/5	10	3.2	E-4	10	B-5	10
Math Applications	40	5	8	2.1	E-4	6	B-5	4
Total Mathematics	112	18/3 32/4 62/5	26	2.5	E-4	4	B-5	4
Stanford Achievement Test, Intermediate Level II (Form A, 1973)								
Vocabularly	50	4	13	3.4	E-5	8	E-6	4
Reading Comprehension	71	4	18	3.3	E-5	11	E-6	6
Word Study Skills	50	25/4 25/5	11	3.0⁻	E-5	10	E-6	8
Total Reading	121	96/4 25/5	29	3.0⁻	E-5	8	E-6	6
Math Concepts	35	4	9	3.5	E-5	6	E-6	4
Math Computation	45	24/3 21/5	12	3.9	E-5	10	E-6	6
Math Applications	40	5	8	3.0⁻	E-5	8	E-6	4
Total Mathematics	120	24/3 35/4 61/5	29	3.2	E-5	4	E-6	2

Stanford Diagnostic Mathematics Test, Red Level (Form A, 1976)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Number System and Numeration	30	3/4 27/5	6	K	B-2	6	E-3	1⁻
Computation	33	5	7	1.2	B-2	7	E-3	1⁻
Applications	30	5	6	1.0	B-2	3	E-3	1⁻
Mathematics Total	93	3/4 90/5	19	1.4	B-2	3	E-3	1⁻

Stanford Diagnostic Mathematics Test, Green Level (Form A, 1976)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Number System and Numeration	36	5	7	2.6	E-3	8	B-6	1
Computation	48	5	10	2.5	E-3	5	B-6	1
Applications	30	5	6	2.6	E-3	14	B-6	1
Mathematics Total	114	5	23	2.5	E-3	5	B-6	1

Stanford Diagnostic Mathematics Test, Brown Level (Form A, 1976)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Number System and Numeration	36	5	7	3.3	E-5	7	E-6	3
Computation	48	5	10	3.2	E-5	9	E-6	3
Applications	33	5	7	2.8	E-5	9	E-6	3
Mathematics Total	114	5	23	3.0	E-5	5	E-6	2

Stanford Diagnostic Reading Test, Red Level (Form A, 1976)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Auditory Vocabulary	36	3	12	1.3	B-2	3	B-3	1
Auditory Discrimination	40	2	20	K	B-2	5	B-3	3⁻
Phonetic Analysis	40	4	10	K	B-2	2	B-3	1⁻
Word Reading	42	3	14	1.1	B-2	13	B-3	3
Reading Comprehension	48	3	16	K	B-2	12	B-3	4
Reading Comprehension Total	90	3	30	1.4	B-2	11	B-3	3

Stanford Disgnostic Reading Test Green Level (Form A, 1976)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Auditory Vocabulary	40	3	13	2.6	E-2	9	B-5	2
Auditory Discrimination	36	12/2 24/3	14	1.4	E-2	11	B-5	4
Phonetic Analysis	36	3	12	2.0	E-2	12	B-5	8
Structural Analysis	60	3	20	K	E-2	8	B-5	2
Literal Comprehension	30	4	8	K	E-2	18	B-5	2
Interential Comprehension	30	24/3 6/4	10	K	E-2	25	B-5	3
Reading Comprehension Total	60	24/3 36/4	17	2.5	E-2	19	B-5	2

Stanford Diagnostic Reading Test, Brown Level (Form A, 1976)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Auditory Vocabularly	40	3	13	3.3	E-4	14	E-6	3
Literal Comprehension	30	4	8	2.9	E-4	21	E-6	10
Inferential Comprehension	30	4	8	2.6	E-4	25	E-6	10
Reading Comprehension Total	60	4	15	2.9	E-4	18	E-6	7
Phonetic Analysis	36	3	12	2.4	E-4	12	E-6	14
Structural Analysis	54	24/V* 30/4	13	K	E-4	1	E-6	1
Reading Rate	34	3	11**	4.2**	E-4	42**	E-6	27**

* Number of possible choices varied with each item in this section.

** Rapidly timed subtest which tests the reading rate (number of items completed) of each student. A chance score is not meaningful for such tests.

Sequential Tests of Educational Progress III, Circus Primary C (Form X, 1979)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Reading	45	4	11	1.5⁻	B-2	2		
Mathematics	40	4	10	1.5⁻	B-2	2		

Sequential Tests of Educational Progress III, Circus Primary D (Form X, 1979)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Reading	42	41/4 1/3	11	2.5⁻	E-2	5	B-3	3
Mathematics	40	4	10	2.5⁻	E-2	3	B-3	2

Sequential Tests of Educational Progress III, Intermediate E (Form X, 1979)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Reading	50	4	13	2.9⁻	E-3	5	B-4	5
Vocabulary	30	4	8	2.9⁻	E-3	9	B-4	7
Math Computation	55	4	14	2.9⁻	E-3	3	B-4	3
Math Basic Concepts	50	4	13	2.9⁻	E-3	4	B-4	4

Sequential Tests of Educational Progress III, Intermediate F (Form X, 1979)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Reading	50	4	13	3.9⁻	E-4	4	B-5	3
Vocabulary	30	4	8	3.9⁻	E-4	6	B-5	6
Math Computation	55	4	14	3.9⁻	E-4	2	B-5	2
Math Basic Concepts	50	4	13	3.9⁻	E-4	3	B-5	3

Sequential Tests of Educational Progress III, Circus Primary C (Form X, 1979)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Reading	45	4	11	1.5⁻	B-2	2		
Mathematics	40	4	10	1.5⁻	B-2	2		

Sequential Tests of Educational Progress III, Circus Primary D (Form X, 1979)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Reading	42	41/4 1/3	11	2.5⁻	E-2	5	B-3	3
Mathematics	40	4	10	2.5⁻	E-2	3	B-3	2

Sequential Tests of Educational Progress III, Intermediate E (Form X, 1979)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Reading	50	4	13	2.9⁻	E-3	5	B-4	5
Vocabulary	30	4	8	2.9⁻	E-3	9	B-4	7
Math Computation	55	4	14	2.9⁻	E-3	3	B-4	3
Math Basic Concepts	50	4	13	2.9⁻	E-3	4	B-4	4

Sequential Tests of Educational Progress III, Intermediate F (Form X, 1979)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Reading	50	4	13	3.9⁻	E-4	4	B-5	3
Vocabulary	30	4	8	3.9⁻	E-4	6	B-5	6
Math Computation	55	4	14	3.9⁻	E-4	2	B-5	2
Math Basic Concepts	50	4	13	3.9⁻	E-4	3	B-5	3

Sequential Tests of Educational Progress III, Intermediate G (Form X, 1979)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Reading	50	4	13	4.9⁻	E-5	3	B-6	3
Vocabulary	30	4	8	4.9⁻	E-5	9	B-6	9
Math Computation	55	4	14	4.9⁻	E-5	2	B-6	2
Math Basic Concepts	50	4	13	4.9⁻	E-5	4	B-6	4

Sequential Tests of Educational Progress III, Intermediate H (Form X, 1979)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Reading	50	4	13	5.9⁻	B-6	6	E-6	5
Vocabulary	30	4	8	5.9⁻	B-6	13	E-6	7
Math Computation	55	4	14	5.9⁻	B-6	5	E-6	3
Math Basic Concepts	50	4	13	5.9⁻	B-6	6	E-6	4

Sequential Tests of Educa- [illegible]	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile

Sequential Tests of Educational Progress III (1979) -			
	Circus Primary C	1.5	2.5
	Circus Primary D	2.5	3.5
	Intermediate E	3.5	4.5
	Intermediate F	4.5	5.5
	Intermediate G	5.5	6.5
	Intermediate H	6.5	7.5

APPENDIX B

NAMES OF EXPANDED STANDARD SCORE SCALES
USED BY MAJOR TEST PUBLISHERS

Test	Name of Expanded Standard Score Scale
California Achievement Tests (1977)	Obtained Scale Scores
Comprehensive Tests of Basic Skills (1968, 1973)	Expanded Scale Scores
Gates-MacGinitie Reading Test (1978)	Extended Scale Score
Iowa Tests of Basic Skills (1978)	Standard Score
Metropolitan Achievement Tests (1978)	Scaled Scores
Sequential Tests of Educational Progress III (1979)	Standard Scores
SRA Achievement Series (1978)	Growth Scale Values
Stanford Achievement Test (1973)	Scaled Scores
Stanford Diagnostic Mathematics Test (1976)	Scaled Scores
Stanford Diagnostic Reading Test (1976)	Scaled Scores

APPENDIX C

WORKSHEET FOR CONVERTING OUT-OF-LEVEL TEST SCORES

Used When Publishers Provide Raw Score-to-Percentile Conversion Tables

Test: Stanford Achievement Test, Form A
Subtest: Total Reading
Grade: 8
Date Administered: October 15, 1977

Out-of-level Test Administered: Intermediate II
In-Level Test Recommended by Publisher: Advanced
Expanded Standard Score Scale Name: Scaled Scores

Student Name	Out-of-Level Raw Score	Expanded Standard Score from Out-of-Level Norms	Raw Score from In-Level Norms	Percentile Rank from In-Level Norms
Source of Score	Number Right	Raw Score-to-Scaled Score conversion table, Intermediate II Manual, Part II, pp. 29-30	Raw Score-to-Scaled Score conversion table, Advanced Manual, Part II, pp. 28-29	Scaled Score-to-Percentile Rank conversion table, Advanced Manual, Part II, p. 25 (be sure to use percentile rank tables for grade and time of year test administered)
	EXAMPLE			
Sample 1	83	175	60	70
Sample 2	54	155	41	16
Sample 3	65	162	47.5 (midpoint of interval 47-48)	24.3 (obtained by interpolation)

WORKSHEET FOR CONVERTING OUT-OF-LEVEL TEST SCORES

Used When Publishers Provide Raw Score-to-Percentile Conversion Tables

Test:
Subtest:
Grade:
Date Administered:

Out-of-Level Test Administered:
In-Level Test Recommended by Publisher:
Expanded Standard Score Scale Name:

Student Name	Out-of-Level Raw Score	Expanded Standard Score from Out-of-Level Norms	Raw Score from In-Level Norms	Percentile Rank from In-Level Norms

APPENDIX D

WORKSHEET FOR CONVERTING OUT-OF-LEVEL TEST SCORES

Used When Publishers Provide Expanded Standard Score-to-Percentile Conversion Tables

Test: Metropolitan Achievement Test
Subtest: Total Reading
Grade: 5
Date Administered: October 15, 1977

Out-of-Level Test Administered: Elementary, Form F
In-Level Test Recommended by Publisher: Intermediate, Form F
Expanded Standard Score Scale Name: Standard Scores

Student Name	Out-of-Level Raw Score	Expanded Standard Score from Out-of-Level Norms	Percentile Rank from In-Level Norms
Source of Score	Number Right	Raw Score-to-Standard Score conversion table, Teacher's Handbook, Elementary Level, p. 6	Standard Score-to-Percentile Rank conversion table, Teacher's Handbook, Intermediate Level, p. 8 (be sure to use percentile rank tables for grade and time of year test was administered)
EXAMPLE			
Sample 1	45	58	12
Sample 2	60	64	24
Sample 3	67	69	34

WORKSHEET FOR CONVERTING OUT-OF-LEVEL TEST SCORES

Used When Publishers Provide Expanded Standard Score-to-Percentile Conversion Tables

Test:
Subtest:
Grade:

Out-of-Level Test Administered:
In-Level Test Recommended by Publisher:
Expanded Standard Score Scale Name:

Student Name	Out-of-Level Raw Score	Expanded Standard Score from Out-of-Level Norms	Percentile Rank from In-Level Norms

"END OF DOCUMENT"